ABSTRACT

A system and method that facilitate modeling unobserved speech dynamics based upon a hidden dynamic speech model in the form of segmental switching state space model that employs model parameters including those describing the unobserved speech dynamics and those describing the relationship between the unobserved speech dynamic vector and the observed acoustic feature vector is provided. The model parameters are modified based, at least in part, upon, a variational learning technique. In accordance with an aspect of the present invention, novel and powerful variational expectation maximization (EM) algorithm(s) for the segmental switching state space models used in speech applications, which are capable of capturing key internal (or hidden) dynamics of natural speech production, are provided. For example, modification of model parameters can be based upon an approximate mixture of Gaussian (MOG) posterior and/or based upon an approximate hidden Markov model (HMM) posterior using a variational technique.